DATA SCIENCE

INTRODUCTION

SUNIL GHIMIRE

TABLE OF CONTENTS

01 PREREQUISITE

Things you should know beforehand

PROGRAMMING

01 **DSA**
Data Structures and Algorithms

02 **Languages**
Python / R

03 **Database**
Sql Scripting

04 **Version Control**
Git and Github

05 **Linux**
Linux Commands

MATHEMATICS

01 **Linear Algebra**

Vectors and Matrices

02 **Calculus**

Limit, Derivative and Integration

03 **Probability**

Hypothesis and Testing

04 **Statistics**

Mean, Median, Mode, Std

02 PIPELINE

Discrete steps to progress (CRISP-DM)

BASIC ILLUSTRATION

BUSINESS UNDERSTANDING

BUSINESS OBJECTIVE

Gain insights of business problems, goals and resources for data mining

ASSESS SITUATION

Risk analysis
Cost-Benefit analysis
Requirements & Availability

DATA MINING GOALS

Transfer Business objective to data mining perspective

PROJECT PLAN

Construct a step-by-step blueprint of the project

DATA UNDERSTANDING

DATA COLLECTION

Select most promising attributes only

DATA DESCRIPTION

Focus on quantity and quality of data

DATA EXPLORATION

Data visualization using tables, charts & other tools

QUALITY VERIFICATION

Check for missing data, out of order and other errors

DATA PREPARATION

SELECT DATA

Select which dataset to work with and why

CLEAN DATA

Correct, impute or remove useless data

CONSTRUCT DATA

Derive new attribute from existing ones

INTEGRATE DATA

Combine data from multiple sources

FORMAT DATA

Re-format data as necessary

MODELLING

ALGORITHM SELECTION

Choose an algorithm that works better for the problem

TRAIN TEST SPLIT

Split data into train & test sets (sometimes valid too)

MODEL TRAINING

Implementing Machine Learning model for real

Assess Model

Interpret model based on domain knowledge using test set

EVALUATION

RESULT EVALUATION

Testing the result on some unseen data and see how our model performs on it

REVIEW

Check if the result works good enough i.e. if it meets the threshold values

RETRAIN OR END?

End the process if threshold values are met else retrain

03 CAREER

How are data scientists doing all around?

CAREER

DATA MINING

Collect useful data
from different sources

BUSINESS INTELLIGENCE

Write queries to
generate reports

DATA ANALYST

Runs queries to find
important trends

DATA ENGINEERING

Prepare data to fit to
ML models

MACHINE LEARNING

Build AI models for
the business

04 NETWORKING

Answering questions raised by audience

THANK YOU

Linkedin: ghimiresunil